Aalto University
School of Business

Is ESG investing brave enough to ride a rollercoaster?

An Experiment on Mutual Fund Flows and Sustainability Ratings during a pandemic

Bachelor's Thesis
Oskari Kouti
Aalto University School of Business
Finance
Fall 2020

Aalto University
School of Business

Aalto University, P.O. BOX 11000, 00076
AALTO
www.aalto.fi

Abstract of bachelor's thesis

Author Oskari Kouti

Title of thesis Is ESG investing brave enough to ride a rollercoaster?

Degree Bachelor's Degree

Degree programme Finance

Thesis advisor(s) Theresa Spickers

Year of approval 2020 **Number of pages** 21 **Language** English

Abstract

By examining mutual fund flows in the U.S. before and after the Covid-19 outbreak, I find evidence that mutual fund investors collectively and consistently value sustainability, even during times of significant market volatility. While not definitive, it also seems like this effect is further amplified during the initial three months of the pandemic. During this time, a total of $10.5 billion escaped the Low sustainability funds and $1.6 billion entered highly ranked funds when compared to the average categories. Possible explanations for this phenomenon can be found from preferences of institutional investors, nonpecuniary motives, or naive expectations of higher performance. However, rational performance expectations cannot be justified based on either my data or prior research.

Keywords Fund Flow, Sustainability, ESG, Covid-19

Contents

1. Introduction

Sustainability and societal impact have gained a lot of attention in recent years, as ESG factors, which stands for Environmental, Social, and Governance, are increasingly being used to make business and investment decisions. In addition, sustainable investing has seen explosive growth in recent years with an increase of 42% from the year 2018 to 2019 (Vinet and Zhedanov, 2011). Therefore, it is important to understand to what extent investors value these traits. Whether companies should aim to purely maximize profits and generate value for shareholders has been challenged more and more in recent years, both inside the businesses themselves but also among investors. Some investors believe that such factors are value-adding, while some do not. While the exact motives of individual investors are difficult to pinpoint and measure, we can observe them indirectly through the mutual fund universe for us to better understand what they value and want.

In this paper, I show evidence that mutual fund investors in the U.S. collectively put a positive value on sustainability, and that their preferences are not prone to change even during significant amounts of market volatility. I carry out a study similar to the one of (Hartzmark and Sussman, 2019), in which they measure the changes in sustainability ratings after a shock experienced by mutual funds and expand on it on two fronts. First, by analyzing the trend past the initial launch and "honeymoon" period of the ratings which Hartzmark and Sussman expected to fade away, and second, by observing the effects of a large exogenous shock to the global economy caused by Covid-19 to the aforementioned factors and their robustness. This is especially interesting, as the attention that ESG factors are getting is a rapidly increasing phenomenon. In addition, these factors with their current popularity were yet to experience a big market crash before Covid-19. While there is a distinct difference between the exogenous shock I'm observing and the one researched by Hartzmark and Sussman in that this time fundamentals also change, the core idea of observing the salience of a characteristic remains the same.

Covid-19 is a disease caused by a virus named SARS-CoV-2 that was initially identified in China around the end of 2019 and the start of 2020. Due to its highly infectious nature, it quickly started to spread across the globe. As a result, a massive supply & later a demand shock shook the globe, leading to significant uncertainty on the markets. March 11th, 2020 the World Health Organization WHO declared Covid-19 as a pandemic. While there was small stock market volatility when the virus initially started to spread, the stock markets did not immediately react as the severity of the situation was not yet understood. Since there is not an exact time when the effects of the virus started, I use the S&P 500 Index as a proxy. February 24th, which I use as the starting point, the SP500 recorded a fall of -3.4%, after which it gradually fell to its lowest point in 2020 (as of 15.11.)

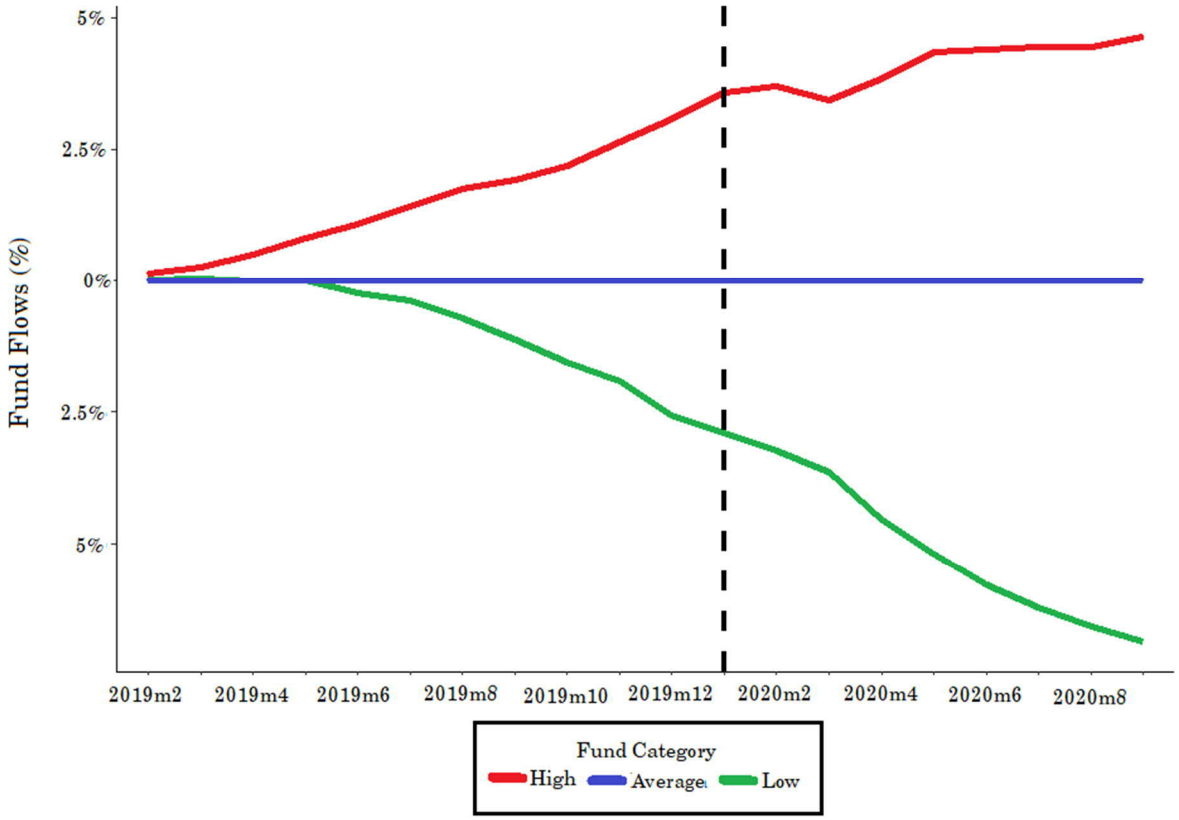
on the 23rd of March, recording a fall of over -35% from its peak (Datastream, 2020). Since the data is monthly, January is considered ex-ante, and February is considered ex-post. Furthermore, despite the pandemic still running strong at the time of writing this thesis, I focus primarily on the first couple of months when the volatility was at its highest.

My time frame of interest includes 12 months before the event up until September of 2020. I explore an event after which mutual funds with assets over $7.5 trillion experienced a large shock of total returns around -24%, with net asset value dropping by around $1.6 trillion. The time frame of interest includes 12 months before Covid-19 up until September 2020, which was the latest data point available when gathering the data. In determining the value that investors put on sustainability, I use the Morningstar Sustainability Ratings, first introduced in 2016, to determine the sustainability of each fund. The Rating system has five categories, from one to five, where the top 10% of funds are ranked as High sustainability and given Five Globes, whereas the bottom 10% are given a Low ranking and One Globe.

The main findings of my paper are illustrated in Figure 1. Consistent with the results of (Hartzmark and Sussman, 2019), mutual fund investors value sustainability as a positive fund attribute, allocating more money into funds with a high sustainability rating, and taking more money out of funds with a low rating. Before taking year-month fixed effects into account, funds rated One Globe experienced outflows of 14.2%, while the funds rated Five Globes only experienced outflows of 2.7% during the observation period of 20 months. Additionally, there is a change in investor behavior during February and March when the volatility was at its highest. Within those months, the differences in inflows and outflows were amplified. During the observation period, overall, all fund classes experienced cumulative net outflows, but I have controlled for time fixed effects to account for that. Like the results of Hartzmark and Sussman (2019), my regressions also resulted in significant results mainly for the Two extremes of One and Five Globes, I've emphasized those in my results. Investors have also been found to pay more attention to discrete ratings rather than continuous ones (Feenberg, Ganguli, Gaule, and Gruber, 2017)

Figure 1: Cumulative Fund Flows (%) by sustainability rating

This figure illustrates the cumulative fund flows from February 2019 to September 2020. High and Low represent Morningstar Sustainability ratings of Five and One Globes respectively, while Average represents the three middle categories. Average is set as the intercept, and year-by-month fixed effects have been controlled for. The dashed line represents January of 2020, which is the month prior to the effects of Covid-19 hit the stock markets.



My paper contributes to the literature in two ways. First, by investigating the differences in fund flows before and after a large exogenous market crash. while the effects of ESG factors during the Covid-19 pandemic have been tested on stocks and ETF's (see (Demers *et al.*, 2020; Folger-Laronde *et al.*, 2020), they have not been tested on Mutual Funds or analyzed based on a difference-in-difference approach. This is motivated by two things: the last big market crash of the financial crisis happening during a time when ESG wasn't so popular, and the fact that the discrete Morningstar Sustainability Ratings have only experienced a bull market before this. The second way in which my paper contributes is by providing an update on the popularity of Morningstar Sustainability Ratings as a driver of mutual fund investments. The first study on the matter by Hartzmark and Sussman (2019) was done on the "honeymoon" period of the ratings, so observing its persistence tells us about its potential long-term effect.

2. Sustainability Ratings

In 2016, Morningstar released their Morningstar Sustainability Ratings, intending to provide investors a discrete, reliable, and objective way to evaluate the ESG factors of portfolios. They are expressed using a rating of One to five Globes, where One Globe stands for Low sustainability and Five Globes stand for High sustainability. The exact breakpoints are described in Table 1. To evaluate the portfolios, Morningstar uses ratings from Sustainalytics that measure the ESG risk of companies, a company which they acquired in May 2020. The rating is a result of a three-step process. First, a Morningstar Sustainability Score is calculated for portfolios reported in the last 12 months. Second, these scores are used to calculate the historic score of the portfolio. Third, a Morningstar Sustainability Rating is assigned based on the portfolio's Historical Portfolio Score in its Morningstar Global Category.

Table 1: Morningstar Sustainability Ratings

Portfolio Rating	Distribution	Descriptive Rank
Five Globes	Highest 10%	High
Four Globes	Next 22.5%	Above Average
Three Globes	Next 35%	Average
Two Globes	Next 22.5%	Below Average
One Globe	Lowest 10%	Low

In November 2019, the measure was changed from using the company ESG Rating to its ESG Risk rating, and buffers between rating increments to add stability were added. This led to discontinuity in the scores, making comparisons between the periods difficult. However, according to Hartzmark & Sussman (2019), the probability of funds staying in the same category from month to month is around 80%. Thus, I have decided to use the same Historical Sustainability Score of portfolios for the observation period when used as a control variable.

3. Data Sources and Summary statistics

All the mutual fund data is from Morningstar Direct. My sample consists of all open-end equity mutual funds domiciled in the US with a Morningstar Sustainability Rating and fund flow data available, and the data is monthly. The start of my observation period is February 2019, and the end is September of 2020. As a result, the time range for the pre-period is 12 months, and for post-Covid-19 it is 8 months, leaving me with up to 20 time-series data points per fund. The sample consists of 12005 individual funds, and 3613 portfolios that are identifiable with a fund-specific ID and thus could be combined with other data sources. The initial data is measured at the share class level, but I perform my analysis at the fund level to avoid double counting. Fund Size (TNA) and dollar flows are calculated as the sum across share classes, while returns, fees, inception date, and Morningstar rating are the average.

Monthly fund flows, the main variable of interest, is measured by Comprehensive Net Fund Flows divided by the Total Net assets of the previous month. I also winsorize the flow variable at the 1% level to get rid of outliers that were initially driving the results. To be consistent, I also winsorize other continuous variables.

Tables 2, 3, and 4 report the summary statistics for the whole sample, the pre-period and post Covid-19 outbreak, respectively. The first thing to note is the trend of high sustainability funds having systematically smaller outflows and higher inflows across all periods, with an average monthly spread of 0.96% before the outbreak and 1.04% after. This solidifies the hypothesis that sustainability is indeed a metric investors value. One and Five Globe funds are the smallest in terms of net assets, which may be explained by funds at extreme ends being less diversified. Additionally, Two globe funds are by far the largest which is potentially explained by the fact that the institutional ownership of the funds in that category is over 80%, which is around 30% higher than it is for other classes. For this effect, I could not find an explanation. Surprisingly, there also seems to be a trend of Sustainable Funds performing better on Morningstar Ratings, but Excess returns show no immediate trend across either period.

Table 2: Summary statistics

This table presents summary statistics for the data. Panel A examines the entire sample period from February 2019 until September 2020. Panel B examines the pre-period before the Covid-19 outbreak until January 2020, and Panel C examines the period during Covid-19 of February 2020 up until September 2020. Statistics are the mean, time-series data is monthly at the fund level, and observations include time-series data.

Panel A: Whole sample

	Fund Flows	Net assets (m)	Age	Morningstar Rating	Excess return	Management Fee	Total observations
One Globe	-0.84%	1555	167	2.69	-0.31%	0.78%	6221
Two Globes	-0.49%	2530	164	3.00	0.12%	0.65%	16715
Three Globes	-0.49%	1871	168	3.20	-0.44%	0.66%	27103
Four Globes	-0.31%	1869	172	3.36	0.05%	0.71%	14715
Five Globes	0.15%	1521	159	3.42	-0.26%	0.75%	5850
Total	0.43%	1970	167	3.16	-0.22%	0.69%	70604

Panel B: Before the Covid-19 outbreak

One Globe	-0.73%	1578	166	2.688	0.13%	0.79%	3839
Two Globes	-0.35%	2534	163	2.997	0.47%	0.65%	10224
Three Globes	-0.39%	1869	167	3.199	0.12%	0.66%	16516
Four Globes	-0.20%	1836	171	3.363	0.43%	0.71%	8975
Five Globes	0.23%	1505	159	3.424	0.23%	0.75%	3544
Total	-0.32%	1964	166	3.162	0.23%	0.69%	43098

Panel C: After Covid-19 breakout

One Globe	-1.01%	1518	168	2.689	-1.03%	0.78%	2382
Two Globes	-0.71%	2525	166	2.997	-0.47%	0.65%	6491
Three Globes	-0.64%	1874	169	3.200	-1.35%	0.66%	10587
Four Globes	-0.47%	1921	173	3.363	-0.57%	0.71%	5740
Five Globes	0.03%	1550	160	3.424	-0.57%	0.75%	2306
Total	-0.60%	1980	168	3.163	-0.96%	0.85%	27506

4. Methodology and Findings

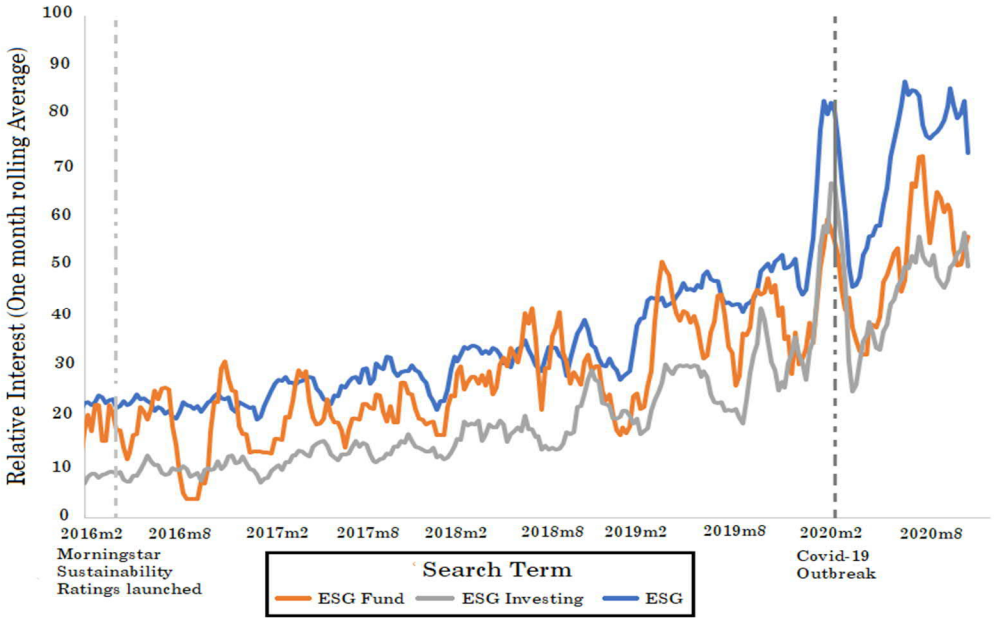
A. Base Results

Do investors still believe in sustainability, and how well did factors related to ESG hold against a large shock to the global economy? To answer these questions, for my base results I start by examining the popularity of ESG related terminology from Google Trends. I then examine fund flows before and after the Covid-19 outbreak by doing panel regressions on various factors and control variables. Furthermore, I look at how each of the Morningstar Sustainability Ratings has held up and to what extent. Lastly, I do a Difference-in-Difference analysis on the two measurement periods. First on an aggregate level, and then by examining individual months.

To determine whether there indeed has been a general increase in interest in sustainability, I initially use Google search trends related to sustainability as a proxy. I observe a period from the start of 2016 up until the September of 2019 to gain a better understanding of what the trend has been since the Morningstar Sustainability Ratings were first introduced, and the trend is clear. Interestingly, the search terms "ESG" and "ESG Investing" both saw their Relative Interest peaks exactly on the week starting on 24.2.2020 when the markets started to take a nosedive, solidifying the hypothesis that ESG factors were in the minds of investors at the start of the pandemic.

Figure 2: Relative Interest in Google Trends for ESG

This graph displays the four-week moving average of the relative interest towards ESG, ESG Funds, and ESG Investing. The lighter dashed vertical line on the left displays the initial launch of Morningstar Sustainability ratings, whereas the dashed vertical line on the right shows the time when the markets first reacted to Covid-19.



In determining the meaningfulness of Morningstar Sustainability Ratings for investors when making investment decisions, examining fund flows provides a great opportunity to answer this question. Unlike individual stocks, the money flowing in and out of funds is not necessarily reflected immediately on the price as the supply is not fixed in the short run. If investors value sustainability and Morningstar Sustainability Ratings, the funds rated with Five Globes will see relatively higher inflows than the funds rated One Globe.

In Tables 3 and 4, I show evidence that investors have given a positive weight to funds with high sustainability ratings, and a negative weight to low sustainability funds. I also include several control variables motivated by prior research to account for potential omitted variable biases. I also include monthly fixed effects to control for the variability across time. In column (1) I examine the dummy variables of the Globe ratings, with Three Globes as the intercept. Additionally, the sustainability Score and the Rank in Global Morningstar Category are included. Whereas both measures were significant, their standalone impact was negligible. One globe funds experienced outflows of almost 0.5%, whereas funds rated Five Globes experienced inflows of 0.77% greater than the Three Globe funds, both of which were extremely significant. This gives a total difference between the extremes of 1.27% with a p-value of less than 0.0001.

The control variables are ones that have been previously shown to affect fund flows. For example, (Ivković and Weisbenner, 2009) found that higher expense ratios lead to a negative fee-fund-flow relationship. Additionally, Morningstar Star Ratings have also shown a positive relationship for fund flows (Guercio and Tkac, 2008), while also being an indirect measure of long-term performance which has also been shown to have a positive flow effect by Ivković and Weisbenner and (Sirri and Tufano, 1998). To further adjust to past performance affecting the returns, Carhart (1997) four-factor alphas and lagged returns have been used. Lastly, Huang, Wei, and Yan (2007) report older funds having lower inflows. Each of these variables had significant results, but their overall effect still did not explain all the flows.

Table 3: Pre Covid-19 Fund Flow Panel Regressions on Sustainability Ratings

This table describes the extent to which Morningstar Sustainability Ratings explain the inflows and outflows of funds. The dependent variable is Fund flows in percentages, and my main explanatory variables are the dummy variables of Globe Ratings. In column (1) I include the raw sustainability scores and ranks in the global category. In column (2) I only include the Globe ratings and in column (3) I omit all the middle ratings. In columns (4) and (5) I include the lagged returns from 1, 3, and 6 months, the log of fund size and age, the management fee, Morningstar Star Ratings, institutional ownership, the Carhart 4-factor model, and the portfolio's sensitivity to economic cycles. All columns include Morningstar Category & time-fixed effects. The data is from February of 2019 to January of 2020, and analysis is done at the fund level. T-stats are in parentheses, and *, ** and *** indicate significance at the 10%, 5% and 1%, respectively.

	(1)	(2)	(3)	(4)	(5)
One Globe	-0.495***	-0.340***	-0.216***	-0.194**	-0.246***
	(-4.33)	(-4.33)	(-2.79)	(-2.30)	(-3.08)
Two Globes	-0.101	0.049		0.100*	
	(-1.34)	(0.889)		(1.74)	
Four Globes	0.317***	0.191***		0.097	
	(4.21)	(3.32)		(0.060)	
Five Globes	0.772***	0.624***	0.387***	0.463***	0.410***
	(6.83)	(7.69)	(4.88)	(5.37)	(5.00)
Difference: Five Globes – One Globe	1.267	0.964	0.603	0.657	0.656
p-value: five globes = one globe	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001
Other Controls	No	No	No	Yes	Yes
Observations	43,086	43,098	43,086	39,382	39,382
R^2	0.006	0.004	0.006	0.016	0.016

The insignificance of the middle globe results points towards the assumption that investors indeed prefer extremes, and introducing controls further amplifies this effect. However, as seen from columns (3) and (5), omitting the middle ratings does not increase the effects of the extremes, and the difference between Five and One globes of 0.657 and 0.656 in columns (4) and (5) is almost the same. Category Percentage ranks and raw sustainability scores were also

included, and those remained insignificant, suggesting that it is the ratings themselves that explain the flows.

Adding more controls reduces the effect of the globes, but they remain sizable with monthly outflows of -0.19% with a t-statistic of -2.3 and inflows of 0.46% with a t-statistic of 5.37. Interestingly, management fees and lagged returns from 1 and 3 months prior together had significant and very large effects, with around the same explanatory power as the extreme globes of One and Five. This is also consistent with the observations of Ivkovic and Weisbenner (2009). Therefore, we can reject the hypothesis that investors were indifferent to measures of sustainability before Covid-19.

Table 4 shows the results after the markets had reacted to Covid-19. Explanatory variables are the same, as are the columns. As the first glaring difference between tables 3 and 4 is that the spread in flows is considerably larger post-Covid-19 with control variables. In column (4) of table 4, the spread is 0.893%, whereas for table 3 the equivalent is only 0.657. This acts as evidence that the crisis has amplified the differences. It is also interesting to note that the outflows and their significance levels have both decreased relatively, as for column (4) the outflows were at -0.2% and significant only at the 10% level, whereas the inflows were significantly higher at 0.69% with a t-stat of 5.71. This indicates that either the interest towards sustainable funds has increased relatively more than it has reduced for low sustainability funds, or that the control variables capture a bigger portion of the flows for One Globe funds. However, the impact of control variables on the explanatory power of Globe ratings has also decreased as they remain relatively stable throughout the columns.

Table 4: Post Covid-19 Fund Flow Panel Regressions on Sustainability Ratings

This table describes the extent to which Morningstar Sustainability Ratings explain the inflows and outflows of funds. The dependent variable is Fund flows in percentages, and my main explanatory variables are the dummy variables of Globe Ratings. In column (2) I only include the Globe ratings and in column (3) I omit all the middle ratings. In columns (4) and (5) I include the lagged returns from 1, 3, and 6 months, the log of fund size and age, the management fee, Morningstar Star Ratings, institutional ownership, the Carhart 4-factor model, and the portfolio's sensitivity to economic cycles. All columns include Morningstar Category & time-fixed effects. The data is from February of 2020 to September of 2020, and analysis is done at the fund level. T-stats are in parentheses, and *, ** and *** indicate significance at the 10%, 5% and 1%, respectively.

	(1)	(2)	(3)	(4)	(5)
One Globe	-0.340**	-0.371***	-0.331***	-0.200*	-0.198**
	(-2.015)	(-3.198)	(-2.882)	(-1.671)	(-2.155)
Two Globes	-0.066	-0.074		-0.063	
	(-0.599)	(0.911)		(-0.780)	
Four Globes	0.167	0.174**		0.123	
	(1.517)	(2.073)		(1.444)	
Five Globes	0.643***	0.672***	0.589***	0.693***	0.660***
	(3.894)	(5.709)	(5.102)	(5.708)	(5.730)
Difference: Five globes – One globe	0.983	1.043	0.920	0.893	0.858
p-value: five globes = one globe	<0.0001	<0.0001	<0.0001	<0.0001	<0.0001
Other Controls	No	No	No	Yes	Yes
Observations	27,500	27,506	27,500	27,469	27,469
R^2	0.003	0.003	0.003	0.021	0.021

One concern is that different industries and funds inevitably have varying characteristics affecting the way they respond to economic shocks, and thus the results would be driven by industry tilts within the funds. However, in their study, (Białkowski and Starks, 2016) found that for their sample the only difference between funds with varying ratings of social responsibility was size, age, and turnover, which indicates that doing well on ESG factors isn't necessarily about firm characteristics. Also, even when controlling for the weighted differences in sensitivity to economic cycles provided by Morningstar, the results saw no noticeable changes. Based on these

results, we can also reject the hypothesis that investors lost their interest in sustainability as the crisis broke out. This also supports the hypothesis that the differences in flows act as a reaction to the market-wide volatility and that investors in the U.S. collectively value sustainability.

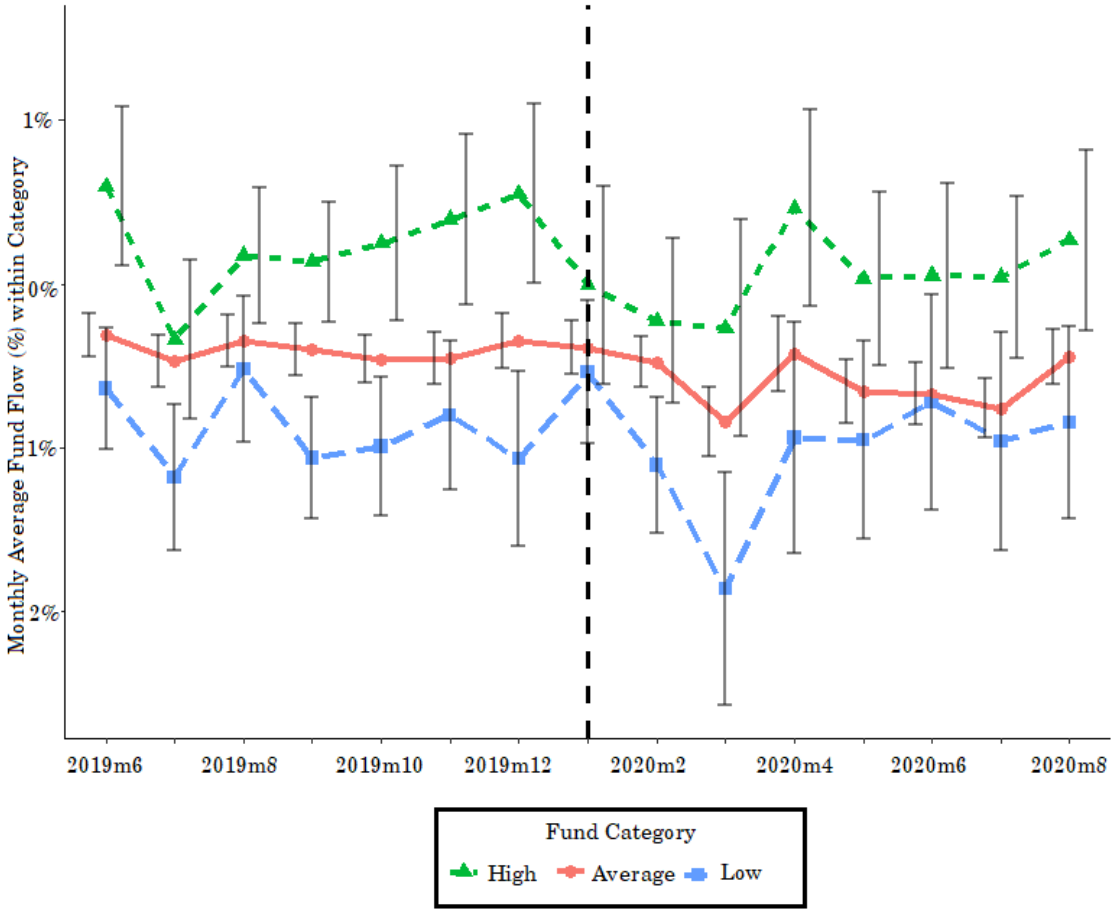
B. Difference-in-Difference analysis

One possibility for the results is that the differences caused within ratings are caused by an unobservable trend or characteristic. To increase the robustness of my results, I conduct a difference-in-difference analysis on the two extreme Globe Ratings while using the Average as a control variable. This allows me to observe the changes in funds with High and Low sustainability between two periods of time and estimate the change attributable to Covid-19. Only comparing the changes within ratings would have the chance of having an unaccounted variable affecting the results. By observing both periods and Globe Ratings together, I will be able to isolate the changes only attributable to the Ratings. I conduct my analysis by first doing a regression on the pre and post period with aggregate data, and then by observing the changes on a month by month basis.

Figure 3 illustrates the monthly average fund flows by category. For the Difference-in-Difference analysis to work, the data must fill the parallel trend assumption. This means that the trends between comparable categories should move in parallel to each other. Generally, this has been the case, with some small occasional hiccups I cannot explain. The vertical bar graphs, centered for Low, slightly to the left for the Average, and slightly to the right for High rated funds, indicate the confidence interval of 95%, for which most of the months have overlap. However, in March of 2020 the Low sustainability fund category, indicated by squares and a blue dashed line, does not overlap with other categories. The dashed vertical line in the middle indicates the period when the markets had not yet reacted.

Figure 3: Difference-in-Difference Summary Data, Monthly Average Fund Flows

This figure illustrates the Monthly average fund flows 6 months prior and 6 months after the Covid-19 outbreak. Categories are based on the Globe ratings, with One globe representing Low sustainability, and Five globes High sustainability. The three categories in the middle are then summed and combined to form the average.



To run the regression, the model I use is based on a regression model by Card (cited in (Angrist and Pischke, 2008), which is displayed in model (1) is used to do the analysis without control variables.

$$Y_{ist} = \gamma_s + \lambda_t + \beta(C_s \cdot d_t) + X'_{ist}\delta + e_{ist} \quad (1)$$

The dependent variable Y_{ist} is fund flows unique to time and category. γ_s stands for the category variables of Low, Average, and High, with average as the control and intercept. λ_t is a variable for different periods, and $\beta(C_s \cdot d_t)$ is the interaction term, where C_s stands for Category dummy and

d_t for the time dummy. I later add the term $X'_{ist}\delta$ to the initial model (1) to create model (2), where the added variable stands for all the different control variables I include.

$$Y_{ist} = \gamma_s + \lambda_t + \beta(C_s \cdot d_t) + X'_{ist}\delta + e_{ist} \quad (2)$$

In Table 5 I show the difference-in-difference regression results for a two-period analysis. According to the difference-in-difference analysis, the effect of having a Low sustainability ranking when the Covid-19 outbreak happened was attributed to -0.08% of the negative outflows, and 0.088% of the positive flows. Introducing controls reduced the outflows attributable to low sustainability to -0.129% per month, and for high sustainability, the flows also decreased to 0.055%. While the general results and their tilt supports the hypothesis, the results are all insignificant. This may be explained by the relatively long timeframe, and means tilting the results, as the highest volatility for returns only happened for two months.

Table 5: Difference-in-Difference Regression with two periods

This table describes the combined estimated change in fund flows (%) from period 1 to 2 attributable to High or Low sustainability Ratings. Low sustainability stands for One Globe and High for Five Globes respectively. The omitted category is Average. Column (1) is regressed with model (1) described above, whereas column (2) is regressed with model (2). Controls included in model (2) are lagged returns from 1, 3 and 6 months, log of fund size and age, management fee, Morningstar Star Ratings, institutional ownership, Carhart four-factor model and the portfolio's sensitivity to economic cycles. The timeframes are from February 2019 to January 2020 for the pre-period, and February 2020 to September 2020 for the post period. T-stats are in parentheses, and *, ** and *** indicate significance at the 10%, 5% and 1%, respectively.

	(1)	(2)
Difference: Post Period, Low Sustainability	-0.008	-0.129
	(-0.129)	(0.130)
Difference: Post Period, High Sustainability	0.088	0.055
	(0.132)	(0.130)

Table 6 explores a similar regression, just with monthly periods to get a better understanding of the events around the outbreak. For robustness's sake, my regression included the entire observation period from February 2019 to September 2020 but results outside table 6 were insignificant and not relevant to the main question. This time, the effects are significantly higher and statistically significant. Flows attributable to Low sustainability were negative thorough the period, with significant results both with and without control variables. During March, the month during the highest volatility, the difference-in-difference between average and Low rated funds was -0.959% without and -1.37% with control variables, and both had significant t-statistics. A

similar effect is also noticeable in February, where the difference with control variables was -0.71% with a t-statistic of -1.80. Highly rated funds also saw positive differences of up to 0.362%, suggesting that a part of the positive flows was attributable to being ranked High sustainability. These results however are not statistically significant. In total, the flows attributable to being ranked Low or High sustainability had a spread of 1.474% in March. To conclude, I reject the hypothesis that the effect of being in the extreme globes of One or Five Globes has not influenced fund flows.

Table 6: Difference-in-Difference Regression with multiple periods

This table describes the combined estimated change in fund flows (%) between periods around the Covid-19 outbreak attributable to High or Low sustainability Ratings. The omitted category is Average. Columns marked (1) are regressed with model (1) described above, whereas columns marked (2) are regressed with model (2). Controls included in model (2) are lagged returns from 1, 3 and 6 months, log of fund size and age, management fee, Morningstar Star Ratings, institutional ownership, Carhart four-factor model, and the portfolio's sensitivity to economic cycles. The timeframes are from January 2020 to April. T-stats are in parentheses, and *, ** and *** indicate significance at the 10%, 5% and 1%, respectively.

	(1) Low	(1) High	(2) Low	(2) High
Difference: 2020m1	-0.088	0.120	-0.226	0.292
	(-0.224)	(0.297)	(-0.575)	(0.745)
Difference: 2020m2	-0.569	-0.012	-0.710*	-0.051
	(-1.453)	(-0.030)	(-1.804)	(-0.131)
Difference: 2020m3	-0.959**	0.311	-1.371***	0.103
	(2.442)	(0.767)	(-3.483)	(0.263)
Difference: 2020m4	-0.449	0.626	-0.624	0.362
	(-1.453)	(1.539)	(-1.585)	(0.925)

5. Why do investors value sustainability?

I now explore three hypotheses based on previous research and my tests as to why investors may value sustainability. My first hypothesis of Rational expected performance of sustainability measures being a predictor of abnormal returns is based on the fundamental financial theory of rational investors and their power to set prices. I then test the hypothesis of institutional behavior contributing to the flows, since between 30-80% of the funds observed were owned by institutional investors. Lastly, motivated by prior research (Hartzmark and Sussman, 2019), I test the hypothesis of naïve performance expectations and nonpecuniary motives.

A. Rational expected Performance

The first hypothesis is based on the rational performance expectations of investors. If sustainability and the return expectations of investors are tied to higher returns, I should find positive coefficients for the Morningstar Sustainability Ratings in the regression analysis. Since the observation period is so short, and the effects are similar for both pre and post periods, I have conducted the analysis on the whole dataset. When observing fund flows in section 4, the lagged returns of one, three, and six months had a combined monthly explanatory power of 0.16% and very significant t-statistics of 5.7, 5.7 and 11.6, respectively.

In table 7, I examine the returns of the portfolios for each of the Globe Ranking based on their monthly excess returns. Overall, the excess returns observed with both regression models are all negative and statistically significant. The spreads between One Globe and Five Globes are 0.41 and 0.16 for CAPM and Carhart, respectively, and for both models, Five Globe funds have performed worse. Different Globes also do not seem to have a trend relative to sustainability, as the middle categories have performed worse. While the period is relatively short and experiences high volatility, the results are also supported by previous research. For example, Auer and Schuhmacher (2016) and Borgers, Derwall, Koedijk, and Horst (2015) also find that funds with high sustainability scores have generally performed worse than their peers.

Table 7: Excess returns, CAPM alphas and Carhart-4 factor alphas

This table describes the monthly Excess returns, CAPM alphas and Carhart 4-factor alphas for each of the globes from the period of February 2019 to September 2020. For the CAPM alpha and Carhart 4-factor model, the dependent variable is monthly excess returns. T-stats are in parentheses, and *, ** and *** indicate significance at the 10%, 5% and 1%, respectively.

	Excess Return	CAPM alpha	Carhart 4-factor alpha
One Globe	-0.31*	-1.08***	-0.38***
	(1.86)	(-9.84)	(-2.64)
Two Globes	0.12***	-1.66**	-0.70***
	(-4.02)	(-25.57)	(-8.23)
Three Globes	-0.44***	-1.51***	-0.71***
	(-3.09)	(-29.94)	(-10.70)
Four Globes	0.05**	-1.15***	-0.42***
	(2.01)	(-16.99)	(-4.76)
Five Globes	-0.26	-1.49	-0.54***
	(-0.96)	(-12.68)	(-3.48)

In table 8, I examine the excess returns on the Morningstar Globe ratings and gradually include control variables. In columns (1) and (2) I only include the Globe Ratings, and for columns (3) and (4) I include the same controls that I did in Table 3 in section 4A. Comparing columns with Globes that include all ratings with columns that only include the extremes and omit the middle categories, we see that the difference grows substantially from the former to the latter. This indicates that the middle categories are of little weight despite the extremely significant t-statistic in column (3) for four globes. As both the tables support the view that the returns have not been better for Five Globe funds than One Globe funds, we can reject the hypothesis that the fund flows would be attributable to rational performance expectations.

This result is further supported by prior research, as Hartzmark and Sussman (2019) found that the results were the same also with value-weighted returns. Additionally, they also compared the excess returns to market and Vanguard benchmarks with methodology from (Berk and van Binsbergen, 2015). Elton, Gruber, and Busse (2004) also found that the cash flows of average investors are not fully reflected on performance. Lastly, the small but significant overperformance of One Globe funds may be explained by the effect of investors preferring to not hold “sin stocks”, which may cause the relatively better performance (Hong and Kacperczyk, 2009). According to the efficient market hypothesis, it is also unlikely that a factor that has recently gained so much attention would not be priced in already.

Table 8: Excess return regressions on Morningstar Globe Ranking

This table describes the extent to which excess returns are explained by Morningstar Globe Ratings. The dependent variable is excess returns in percentages, and my main explanatory variables are the dummy variables of Globe Ratings. In column (1) I only include the Globe ratings and in column (2) I omit all the middle ratings. In columns (3) and (4) I include the lagged returns from 1, 3, and 6 months, the log of fund size and age, the management fee, Morningstar Star Ratings, the Carhart 4-factor model, and the portfolio's sensitivity to economic cycles. All columns include Morningstar Category, historical sustainability score & time-fixed effects. The data is from February of 2019 to September of 2020, and analysis is done at the fund level. T-stats are in parentheses, and *, ** and *** indicate significance at the 10%, 5% and 1%, respectively.

	(1)	(2)	(3)	(4)
One Globe	0.20	0.57***	0.25	0.58***
	(1.20	(4.53)	(1.46)	(4.47)
Two Globes	-0.248**		-0.22	
	(-2.28)		(-1.97)	
Four Globes	0.50***		0.47***	
	(4.55		(4.17)	
Five Globes	0.30*	-0.18	0.22	-0.23*
	(1.824)	(-1.41)	(1.29)	(-1.76)
Difference: Five Globes – One Globe	0.10	-0.75	0.03	-0.81
p-value: five globes = one globe	<0.0001	<0.0001	<0.0001	<0.0001
Other Controls	No	No	Yes	Yes
Observations	69,049	69,049	65,321	65,321
R^2	0.368	0.368	0.377	0.377

B. Institutional behavior

The second hypothesis is that the behavior of institutional actors who face constraints have driven the flows. Due to institutions having both mandates to invest in certain asset classes and types of funds, and managers possibly facing social pressure to avoid sin stocks, their investment choices might not always be driven purely by returns. I test this hypothesis on two metrics. First, I analyze the relative institutional ownership of the portfolios in my sample. Second, I review the literature on the behavior of large institutions.

To determine the impact of institutional ownership on the fund flows, I return to the regressions done in section 4A. In the pre-period regression in Table 3 column (4), the regression coefficient for institutional ownership was 2.49% which is relatively large. However, the t-statistic was only 0.48, leaving the result very insignificant. In the post-Covid-19 regression in table 4 column (4), the coefficient is very large at 6.34%, but again with an insignificant t-statistic of -0.95. When doing similar tests with and without various control variables, the results did not change and therefore based on these tests I cannot reject the hypothesis that the behavior of institutions would be able to explain the flows. When looking at raw fund asset data, the institutional ownership of One Globe funds fell by around 3% during the observation period, while the ownership of Five Globe funds increased from 42% to 48%. However, the effect of these changes has been mostly captured by other variables.

Hartzmark and Sussman (2019) also did a similar exercise, where they tested whether institutional share classes saw different interactions to fund flows than non-institutional ones, and the results were insignificant. Furthermore, they also found the within fund differences in share classes to be insignificant. Cao, Titman, Zhan, and Zhang (2018) also found that socially responsible institutions are less prone to react to mispricing when the trading is against their preferences of ESG performance. A similar finding is also done by Dyck, Lins, Roth, and Wagner (2019), who found that institutional ownership predicted higher environmental and social scores for firms, providing evidence that certain institutions systematically value sustainability. However, the suggested explanation of Hartzmark and Sussman that recently enacted regulations would attribute to the changes is likely not true, as (Boutchkova, Gonzalez, and Zhang, 2020) found that the EU and US showed similar results despite the regulation not being enacted in the US.

While there are signs of institutions having a preference towards high ESG scores and institutional ownership increasing for High Sustainability funds during the observation period, the evidence is not significant enough to account for all the effects. Hartzmark and Sussman

(2019) present two potential reasons for this. First, institutional investors might be behaving similarly to noninstitutional investors either willingly or by coercion. Second, this metric might not be representative of the actions of institutional investors, as individual investors might also be trading on their retirement accounts. Therefore, I cannot reject the hypothesis that institutional behavior would be insignificant.

C. Naïve performance Expectations and Nonpecuniary Motives

The last hypothesis is that investors naïvely expect higher returns from funds with high sustainability, or that they have nonpecuniary motives. As this hypothesis is more difficult to judge based on data from Morningstar or my available resources, I will review previous literature for evidence. According to Hartzmark and Sussman (2019), the nonpecuniary motives hypothesis can be divided into three sections: altruism, warm glow, and social motives. To differentiate between the performance expectation hypothesis and the nonpecuniary hypothesis, they ran an experiment on U.S. MBA Students and MTurk participants on their investment preferences. The results show that sustainability ratings were associated with higher performance and lower risks, despite fundamentals staying the same. They also found that when controlling for risk and performance-based measures, nonpecuniary motives contributed slightly less than half of the difference in flows to One and Five Globe funds for the assumingly more educated investors of MBA students, and slightly more than half for the MTurk participants.

Another way to test expectations is to use Socially Responsible Investing (SRI) as a proxy for sustainability as they are both closely linked and may attract similar investors. (Renneboog, Ter Horst and Zhang, 2011) find that SRI funds are less sensitive to lagged negative returns than conventional funds, which cannot be explained by performance expectations. This finding is further supported by (Bollen, 2007), who finds that SRI funds are also more sensitive to lagged positive returns, implying that SRI investors overreact to good news and underreact to bad news. Furthermore, SRI investment decisions have also been linked to social preferences and social signaling (Riedl and Smeets, 2017).

While there seems to be reason to believe that both naïve performance expectations and nonpecuniary motives contribute to the flows, Boutchkova et al. (2020) take a different approach for explaining the altruistic argument. They study shocks to altruism and find that natural disasters or other such events have a positive relationship with increased philanthropy, but no evidence of changes in levels of philanthropy increased sustainable investing is shown.

According to them, altruism can often be explained by personal gains such as tax deductions, "warm glow" utility or social reasons. Consequently, it is also found to be a driver of philanthropy, but not a driver of sustainable investing, no matter the underlying motives and their "purity". Therefore, it is suggested that altruism and sustainable investing are not connected and thus altruism does not account for the behavior. Although not definitive, there seems to be a good reason to believe that both naïve performance expectations and nonpecuniary motives play a role in explaining the fund flows, and thus the hypothesis of these factors not contributing cannot be rejected.

6. Conclusions

I find evidence that investors in the U.S. collectively put a positive value on sustainability by assessing fund flows before and after the initial market crash Caused by Covid-19 between February and March of 2020, and reject the hypothesis that investors are indifferent to this rating. For One Globe funds, this attributed to higher monthly outflows of around -$950 million before and -$900 million after Covid-19, and for Five Globe funds monthly inflows of around $2 050 million before and $3 105 million after Covid-19. This supports the notion that a significant number of modern investors believe that sustainability is value-adding.

Additionally, I present evidence that the combined effect of belonging to an extreme Morningstar Sustainability Category of One or Five Globes and the shock from the market crash between February and March of 2020 attributed to negative fund outflows for One Globe funds, and to positive fund inflows for Five Globe funds. The economic impact of this combined effect in the first three months of the pandemic was that funds rated One Globe saw outflows of up to -$10.5 billion lower and funds rated Five Globes saw inflows of up to $1.6 billion higher compared to the Two, Three, and Four Globe -categories.

To determine why the mutual fund investors collectively put a positive value on sustainability, I test three hypotheses based on previous research. First, institutional investors might play a role in the aggregate flows, as they own a considerable chunk of the total assets. This is supported by research on the behavior and preferences of institutional investors, but the results are not definitive to reject the hypothesis that they do not play a role. Second, rational performance expectations are likely not the cause, as based on my findings and previous research (Hartzmark and Sussman, 2019), there seems to be an inverse relationship between excess returns or abnormal returns and sustainability, and Five Globe funds have underperformed on average. Lastly, the flows seem to be also attributable to both naïve return expectations and nonpecuniary

motives (Hartzmark and Sussman, Boutchkova et al., 2020). While the exact motives within either option cannot be pinned, altruistic motives are likely not the explanation.

While I have answered most of the questions that may explain this phenomenon, some questions remain unanswered. First, gaining a better understanding of what drives the naïve expectations of higher expected returns would help better explain those motives. In addition, some things such as forward-looking return expectations and measures could try to be eastimated. Historic measures might not be able to explain the phenomena right now, but what about the more distant future? Maybe ESG factors are sleeping giants, and the returns will only realize further down the line. Lastly, understanding the underlying assets of the categories, possible industry tilts and their connectedness would be a question worth answering.

References:

Angrist, J. D. and Pischke, J. S. (2008) *Mostly harmless econometrics: An empiricist's companion*, *Mostly Harmless Econometrics: An Empiricist's Companion*. doi: 10.1111/j.1475-4932.2011.00742.x.

Auer, B. R. and Schuhmacher, F. (2016) 'Do socially (ir)responsible investments pay? New evidence from international ESG data', *Quarterly Review of Economics and Finance*, 59, pp. 51–62. doi: 10.1016/j.qref.2015.07.002.

Berk, J. B. and van Binsbergen, J. H. (2015) 'Measuring skill in the mutual fund industry', *Journal of Financial Economics*, 118(1), pp. 1–20. doi: 10.1016/j.jfineco.2015.05.002.

Białkowski, J. and Starks, L. T. (2016) 'SRI Funds: Investor Demand, Exogenous Shocks and ESG Profiles', *Unpublished working paper*, (11), pp. 1–43. Available at: www.ussif.org.

Bollen, N. P. B. (2007) 'Mutual fund attributes and investor behavior', *Journal of Financial and Quantitative Analysis*, 42(3), pp. 683–708. doi: 10.1017/s0022109000004142.

Borgers, A. *et al.* (2015) 'Do social factors influence investment behavior and performance? Evidence from mutual fund holdings', *Journal of Banking and Finance*, 60, pp. 112–126. doi: 10.1016/j.jbankfin.2015.07.001.

Boutchkova, M., Gonzalez, A. and Zhang, S. (2020) 'Is Sustainable Investing Driven by Altruism: Evidence from Shocks to Philanthropy', *SSRN Electronic Journal*. doi: 10.2139/ssrn.3575000.

Cao, J. *et al.* (2019) 'ESG Preference and Market Efficiency: Evidence from Mispricing and Institutional Trading', *SSRN Electronic Journal*. doi: 10.2139/ssrn.3353623.

Datastream. (2020) *Thomson Reuters Datastream*. [Online]. Available at: Subscription Service (Accessed: November 2020)

Demers, E. *et al.* (2020) 'ESG Didn't Immunize Stocks Against the Covid-19 Market Crash', *SSRN Electronic Journal*. doi: 10.2139/ssrn.3675920.

Dyck, A. *et al.* (2019) 'Do institutional investors drive corporate social responsibility? International evidence', *Journal of Financial Economics*, 131(3), pp. 693–714. doi: 10.1016/j.jfineco.2018.08.013.

Elton, E. J., Gruber, M. J. and Busse, J. A. (2004) 'Are Investors Rational? Choices among Index Funds', *Journal of Finance*, 59(1), pp. 261–288. doi: 10.1111/j.1540-6261.2004.00633.x.

Feenberg, D. *et al.* (2017) 'It's good to be first: Order bias in reading and citing NBER working papers', *Review of Economics and Statistics*, 99(1), pp. 32–39. doi: 10.1162/REST_a_00607.

Folger-Laronde, Z. *et al.* (2020) 'ESG ratings and financial performance of exchange-traded funds during the COVID-19 pandemic', *Journal of Sustainable Finance and Investment*. doi: 10.1080/20430795.2020.1782814.

Guercio, D. Del and Tkac, P. A. (2008) 'Star power: The effect of morningstar ratings on mutual fund flow', *Journal of Financial and Quantitative Analysis*, 43(4), pp. 907–936. doi: 10.1017/S0022109000014393.

Hartzmark, S. M. and Sussman, A. B. (2019) 'Do Investors Value Sustainability? A Natural Experiment Examining Ranking and Fund Flows', *Journal of Finance*, 74(6), pp. 2789–2837.

doi: 10.1111/jofi.12841.

Hong, H. and Kacperczyk, M. (2009) 'The price of sin: The effects of social norms on markets', *Journal of Financial Economics*, 93(1), pp. 15–36. doi: 10.1016/j.jfineco.2008.09.001.

Ivković, Z. and Weisbenner, S. (2009) 'Individual investor mutual fund flows', *Journal of Financial Economics*, 92(2), pp. 223–237. doi: 10.1016/j.jfineco.2008.05.003.

Renneboog, L., Ter Horst, J. and Zhang, C. (2011) 'Is ethical money financially smart? Nonfinancial attributes and money flows of socially responsible investment funds', *Journal of Financial Intermediation*, 20(4), pp. 562–588. doi: 10.1016/j.jfi.2010.12.003.

Riedl, A. and Smeets, P. (2017) 'Why Do Investors Hold Socially Responsible Mutual Funds?', *Journal of Finance*, 72(6), pp. 2505–2550. doi: 10.1111/jofi.12547.

Sirri, E. R. and Tufano, P. (1998) 'Costly search and mutual fund flows', *Journal of Finance*, 53(5), pp. 1589–1622. doi: 10.1111/0022-1082.00066.

Vinet, L. and Zhedanov, A. (2011) *A 'missing' family of classical orthogonal polynomials*, *Journal of Physics A: Mathematical and Theoretical*. doi: 10.1088/1751-8113/44/8/085201.